Matching Bias in Syllogistic Reasoning: Evidence for a Dual-Process Account from Response Times and Confidence Ratings

Edward J. N. Stupple*

University of Derby, UK

Linden J. Ball

Lancaster University, UK

Daniel Ellis

University of Derby, UK

*Corresponding author:
Centre for Psychological Research
University of Derby
Kedleston Road
Derby, DE22 1GB, UK

Email: e.j.n.stupple@derby.ac.uk
Tel: +44 (0) 1332593061

Please note that this is a pre-published version of this paper and that the published article may differ

Abstract

We examined matching bias in syllogistic reasoning by analysing response times, confidence ratings and individual differences. Roberts' (2005) 'negations paradigm' was used to generate conflict between the surface features of problems and the logical status of conclusions. The experiment replicated matching bias effects in conclusion evaluation (Stupple & Waterhouse, 2009), revealing increased processing times for matching/logic 'conflict problems'. Results paralleled chronometric evidence from the belief bias paradigm indicating that logic/belief conflict problems take longer to process than non-conflict problems (Stupple, Ball, Evans, & Kamal-Smith, 2011). Individuals' response times for conflict problems also showed patterns of association with the degree of overall normative responding. Acceptance rates, response times, metacognitive confidence judgements and individual differences all converged in supporting dual-process theory. This is noteworthy because dual-process predictions about heuristic/analytic conflict in syllogistic reasoning generalised from the belief bias paradigm to a situation where matching features of conclusions, rather than beliefs, were set in opposition to logic.

Keywords: Matching bias; Deduction; Dual-process theory; Individual differences; Reasoning

Matching Bias in Syllogistic Reasoning: Evidence for a Dual-Process Account from Response Times and Confidence Ratings

Human reasoning is fallible. Although we are sometimes capable of impressive feats of analysis, we are also susceptible to failures in intuition or inference and display tendencies to be biased by preconceptions and beliefs (Evans, 1989). A popular approach taken to account for this contrast comes in the form of dual-process theories of reasoning and judgment (e.g., Evans, 2007; Gigerenzer & Goldstein, 1996; Stanovich, 2004). In broad terms, such theories dissociate fast and frugal 'snap' judgements from slow, effortful and methodical analyses and claim that such dissociable 'heuristic' or 'analytic' thinking activities are underpinned by separate cognitive systems or processes (see Kahneman, 2011, for a recent overview of key concepts in dual-process theory by one of its main proponents). Beyond a basic, broad-brush distinction between heuristic and analytic thinking, however, a wide variety of dual-process approaches have been developed that often differ in their defining features and in their assumptions regarding the cognitive architecture associated with heuristic and analytic processes (e.g., for reviews see Evans, 2007, 2009, 2011).

Dual-process approaches do not simply account for why we reason effectively on some problems and poorly on others; they also capture the internal *conflicts* that arise in judgement and reasoning tasks, when people's analytic processes clash with their heuristically-based intuitions (e.g., Bonner & Newell, 2010; De Neys, 2006a; De Neys, Cromheeke, & Osman, 2011; Evans & Ball, 2010). Arguably, the quintessential paradigm that has established the presence of heuristic/analytic conflicts concerns the study of *belief bias* in syllogistic reasoning (Evans, Barston, & Pollard, 1983). Syllogisms are deductive arguments that involve two premises and a conclusion, each of which contains one of the following quantifiers: *all, no, some* or *some…not*. The introduction of belief-oriented content into syllogisms can produce a conflict between analytic processes that are directed toward a logical evaluation of the presented conclusion and heuristics processes that are driven by the belief status of the conclusion. For example, the conclusion to the following syllogism is logically *invalid* despite its apparent believability: "All flowers need water. All roses need water. Therefore, all roses are flowers". Belief/logic conflicts are revealed in the processing effort that people devote to particular syllogisms. For example, eye-tracking data have shown that many people spend significantly longer inspecting belief/logic conflict syllogisms than ones where the logic and belief status of conclusions are congruent (Ball, Phillips, Wade, & Quayle, 2006).

The belief bias paradigm also gives rise to a classic pattern of findings relating to conclusion endorsement rates that is again interpretable from a dual-process perspective. First, more valid conclusions are accepted than invalid ones, which is indicative of the presence of some degree of analytic processing. Second, more believable conclusions are accepted than unbelievable ones, which is indicative of the influence of heuristic processing. Third, there is an interaction between validity and believability, such that the effects of beliefs are greater on invalid than valid conclusions. Explaining this interaction effect has led to numerous dual-process accounts that centre on the competition that arises between heuristic and analytic processes for particular problems (for a contrary, single-process view see Dube, Rotello, & Heit, 2010). Reviewing the finer details of contemporary dual-process accounts of belief bias is beyond the scope of this paper, although we will describe key aspects of such theories when they bear upon the interpretation of the findings we report. Suffice it to say for now, however, that one particular theory - the *selective processing model* advanced by Evans (2000) - continues to dominate the belief bias field because of its capacity to explain not only conclusion endorsement rates but also evidence relating to individual differences in processing latencies for different problem types (Stupple, Ball, Evans, & Kamal-Smith, 2011; see also Ball, 2010). The supremacy of this model underscores how important it is for any dual-process theory of reasoning to be able to capture a wide range of convergent data beyond mere conclusion acceptance rates.

Stupple and Waterhouse (2009) have recently demonstrated that conflicts between heuristic and analytic processes not only arise in tasks that invoke belief bias but also occur with abstract syllogistic reasoning problems involving *belief-neutral* contents. Their predictions were derived from dominant theories of deductive reasoning, which can broadly be categorised as ‘surface feature theories’ and ‘analytic theories’. Surface feature theories include the *matching theory* of Wetherick and Gilhooly (1995) and the *probability heuristics model* of Chater and Oaksford (1999). Both theories are based on the principle that the surface features of problems determine responses, with no particular expectations arsing for differential processing times for different problems. In contrast, for theories that emphasise analytic processes such as the *mental models theory* (e.g., Bucciarelli & Johnson-Laird, 1999) and the *mental logic theory* (Rips, 1994), decisions about conclusion validity are primarily made on the basis of mental representations of the underlying logical structure of problems (note, however, that mental models theory includes heuristic processes, particularly in the initial stages of model formation and, as such, can be considered a dual-process theory; see Johnson-Laird, 1983). Unlike surface feature theories, analytic theories make predictions regarding differential processing times for different problem types.

Stupple and Waterhouse (2009) utilised the expanded set of syllogisms devised by Roberts (2005) to provide a novel test of these aforementioned theories. Roberts used negations and double negations to increase the set of syllogistic forms from 64 possible problems to 576. This expanded set of problems allowed Stupple and Waterhouse to test novel hypotheses using problems where the manipulation of surface features (i.e., the use of the double negated premise form ‘No…not’ versus the equivalent affirmative form ‘All’) could conflict with the underlying logic of the problem. Although it could be argued that quantifier combinations involving double negation arise infrequently in normal, everyday communication, they do nevertheless occur occasionally, for example, when someone utters statements such as ‘there are no luxury items that are not desirable’ or ‘there are no fast cars that are not dangerous’. The analysis of how people understand and reason with double negation has a long history in linguistics research (e.g., see Dowty, 1994, for important proposals) and we suggest that any psychological theory purporting to account for deductive reasoning should ideally be able to explain not only how people make inferences from commonplace quantifier formats but also how they deal with less frequent constructions such as double negation.

Stupple and Waterhouse’s (2009) study, however, failed to arbitrate decisively between surface feature and analytic theories of reasoning. Surface feature theories gained support from the high proportion of ‘matching’ responses that were given, whereby participants showed a bias to endorse conclusions that matched the terms used in the premises, irrespective of the logic of the problem. However, surface feature theories could not account for the reliable differences in response times engendered by the use of double negated premises. In contrast, analytic theories offered a ready explanation for the increased response times that arose with double negated premises, but could not easily explain the matching tendency in the response data. Stupple and Waterhouse instead proposed a dual-process account of the data, whereby the dominant processing was analytic, but where, in many cases, this slower processing competes with and is defeated by a matching heuristic. The data were taken as being supportive of either a parallel dual-process account (e.g., Sloman, 2002) or a hybrid dual-process account involving both sequential and parallel components (cf. Evans, 2009; see also Ball, 2010).

The experiment reported in the present paper aimed to extend Stupple and Waterhouse's (2009) findings by manipulating problem complexity and by examining individual differences in reasoning processes as reflected in response times and confidence judgements. Two different levels of problem complexity were introduced by contrasting the syllogistic figure AB-BC against the figure BA-CB, as previously applied in the belief bias paradigm by Stupple and Ball (2008). The AB-BC figure has been shown to be reliably easier to reason with than the BA-CB figure (e.g., Johnson-Laird & Bara, 1984), with Stupple and Ball (2007) presenting evidence for the mental models claim that reasoners can more readily form integrated models of premises when middle terms are contiguous than when they are not (see also Espino, Santamaria, & Garcia-Madruga, 2000). Gilhooly (2005) has suggested that as the complexity of a reasoning task increases then more simplistic reasoning strategies such as matching will be employed to enable the generation of a response under heightened cognitive load. Gilhooly's claim did not previously find support within the double negation paradigm (Stupple & Waterhouse, 2009). In fact, the data showed that the simpler affirmative premises that characterise traditional syllogisms led to significantly higher rates of matching responses than the more cognitively demanding double negated premises. However, the possibility could not be ruled out that participants were simply more reluctant to accept double negated conclusion forms given the observation that fewer of these were accepted compared to traditional affirmative conclusions. In the present experiment we directly manipulated syllogism complexity by using two figural forms: the easier AB-BC form and the harder BA-CB form. Doing this enables a fair test to be conducted of Gilhooly's (2005) prediction from the matching theory, with the expectation being that the more difficult figure should give rise to an increased tendency for participants to accept conclusions that match the surface features of the premises and to reject conclusions that mismatch the surface features.

Stupple and Ball (2008) have argued that converging data deriving from the study of heuristic/analytic conflict problems are essential for corroborating theories of dual processes in reasoning. The problems in the present experiment were, therefore, designed to be either: (1) *conflict problems*, that is, items with valid/non-matching conclusions or invalid/matching conclusions; or (2) *non-conflict problems*, that is, items with valid/matching conclusions or invalid/non-matching conclusions. From the perspective of *parallel* dual-process accounts (e.g., Sloman, 1986; Stupple & Ball, 2008), conflict problems are predicted to display increased response latencies relative to non-conflict problems, since it is assumed that many participants will be sensitive to heuristic/analytic conflicts and will commit additional analytic effort (and therefore increased time) toward the attempted resolution of such conflicts. Thus, in the case of a valid/non-matching conclusion, further analytic processing would be predicted to be deployed in an attempt to falsify the non-matching conclusion. Such falsification would, however, be likely to fail, since the conclusion is, in fact, valid. In the case of an invalid/matching conclusion, further analytic processing would again be engendered, this time in an effort to confirm the matching conclusion. This validation, however, would presumably fail, given that the conclusion is invalid. A parallel-process account would also predict that there should be individual variation in response times to conflict problems, dependent on people's propensity to commit extra analytic processing effort to resolving conflicts. This means that longer response times should generally be associated with increased normative responding, since those individuals who commit more analytic effort toward resolving conflicts (thereby taking longer to come to a decision) should also be more likely to come to a normative response. Note that we use the term 'normative' here to refer to responses consistent with traditional syllogistic logic, but we do so in a purely *descriptive* sense (see Elqayam & Evans, 2011; Stupple & Ball, 2011)

In terms of accounts that involve a stronger allegiance to *sequential* heuristic and analytic processes (e.g., so-called 'default-interventionist' theories; see Evans, 2007), at first glance there seems to be little reason for a prediction of increased latencies for conflict relative to non-conflict items. According to a default-interventionist account, whilst default heuristic responses may be subject to analytic intervention, the probability of such intervention is assumed to be *independent* of the outcome of heuristic processes (Evans, 2007). Since analytic intervention is equally likely for all problem types then one might assume that the timecourse of such analytic intervention would also be equivalent. However, this is not the case, since default-interventionist accounts also embody assumptions regarding the specific nature of the analytic intervention that arises for different problem types as well as the effectiveness or *quality* of reasoning that occurs, which is itself a consequence of different levels of reasoning ability (e.g., Evans, 2007; Stupple et al., 2011). Thus, processing latencies can vary dependent on the problem type (e.g., the matching status of the conclusion), the ability of the reasoner (e.g., whether the individual has have a high or low working memory capacity), and no doubt a range of other factors.

Our aim in this paper is not to attempt to arbitrate between different dual-process accounts, especially in light of the questions surrounding the viability of a strong independence assumption when considering the interplay between heuristic and analytic processing (see Elqayam, 2009, for a particularly important discussion of this issue). Rather, we aim to examine the extent to which data deriving from an examination of heuristic/analytic conflicts in abstract syllogistic reasoning (i.e., where the matching features of conclusions are in opposition to logic) can be shown to support the essential assumptions of a general dual-process framework. One such assumption, for example, is that heuristic/analytic conflict problems should engender increased processing latencies when reasoners resolve conflicts in favour of normative responses, but should engender reduced processing latencies when reasoners resolve conflicts in favour of more superficial heuristic responses.

Heuristic/analytic conflict problems in the reasoning domain have also recently received attention from dual-process theorists who are using neuroimaging techniques to explore the neural correlates of conflict detection and conflict resolution mechanisms. For example, De Neys, Vartanian, and Goel (2008) provided evidence that participants show particular patterns of brain activity when presented with heuristic/analytic conflicts. In particular, the anterior cingulate appears to be active in the face of such conflicts, with conflict resolution being correlated with neural activity in the lateral prefrontal cortex, which is associated with the inhibition of heuristically-based responses (see also Luo, Yuan, Qiu, Zhang, Zhong, & Huai, 2008; Luo, Liu, Stupple, Zhang, Xiao, Jia, Yang, Li, & Zhang, in press, for evidence of belief inhibition from ERP studies of belief bias). There is, moreover, evidence that people exhibit decreased confidence when giving intuitive responses that conflict with normative responses (e.g., De Neys & Franssens, 2009; De Neys et al., 2011). Based on these findings, together with the neuroscientific evidence, De Neys and colleagues have suggested that even though many participants may be unable to resist giving an intuitive response, they are nevertheless more sensitive to normative standards than is often assumed in the literature. The findings from De Neys' laboratory have arisen with both classic judgement and decision-making problems (i.e., conjunction fallacy and base-rate fallacy tasks; De Neys et al., 2011) and standard reasoning tasks that are susceptible to belief bias (De Neys, Moyens, & Vansteenwegen, 2010; De Neys &

Franssens, 2009). In contrast, in the present experiment we use confidence ratings to investigate whether these effects generalise to *belief-neutral* syllogisms.

In relation to confidence measures it is also noteworthy that Shynkaruk and Thompson (2006) have conducted a systematic empirical analysis of confidence ratings in syllogistic reasoning. They demonstrated that confidence was affected by non-logical factors such as belief (with greater confidence for believable and unbelievable items than for neutral items) and the time available for responding (with reduced confidence arising when judgements had to be made within a 10 second limit). Moreover, they found support for Kruger and Dunning’s (1999) claims that participants did not have accurate judgements of their own performance, with poor reasoners overestimating their performance and more able reasoners underestimating it. Finally, Prowse Turner and Thompson (2009) identified three factors that appear to be predictive of people’s confidence judgements for syllogistic conclusions: (1) ‘feelings of rightness’ ratings, based on perceptions of processing complexity; (2) external factors such as time limitations; and (3) individual differences. The most relevant of these factors in relation to the present experiment concerns people’s perceptions relating to processing complexity, such as might arise when a participant is aware of competing responses to a problem, as in the case when the response cued by a matching heuristic conflicts with the response arising from analytic considerations. Such a situation would be expected to reduce confidence. In addition, based on earlier work by Quayle and Ball (2000), it would also be predicted that reduced confidence would arise for syllogisms in the more complex figure (BA-CB) relative to the easier figure (AB-BC).

In sum, response times, confidence ratings and conclusion endorsement data were all utilised in the present experiment to test dual-process theory predictions. In particular, it was predicted that for heuristic/analytic conflict problems the production of *non-normative* responses should correlate negatively with response times if the matching status of conclusions promotes a superficial heuristic response. In other words, the less time that is spent on a conflict problem then the more likely it will be for a non-normative heuristic response to be given (i.e., endorsing an invalid matching conclusion or rejecting a valid non-matching conclusion). In addition, it was predicted that there should be an increase in heuristic responses when confidence is low, based on the assumption that heuristic responses reflect a failure of analytic reasoning (either in terms of quality or amount) that results in the heuristic response winning out. On this basis we would also expect the most complex problems to elicit the lowest confidence ratings (cf. Quayle & Ball, 2000) and the greatest degree of matching-based heuristic responding (Gilhooly, 2005).

Method

Participants

Eighty-five participants were recruited via opportunity sampling. They were predominantly University of Derby undergraduates, but also included a small number of postgraduates and staff. Although we did not collect detailed age data we estimate than all participants were between 18 and 60 years of age, with over 90% of the sample being younger than 40 years of age and a majority of these in the 18 to 24 range. All had a minimum level of educational progression to a Bachelor’s degree. None had prior knowledge of reasoning research as assessed by pre-screening at experiment sign-up.

Design

The study used a repeated measures design. Matching and non-matching problem types were created by combining premise surface features (double negated versus traditional affirmative) and conclusion surface features (double negated versus traditional affirmative). Problems were also manipulated according to figure (AB-BC vs. BA-CB) and validity (valid vs. invalid). The dependent measures were conclusion acceptance rates, response times (from initial problem presentation to an acceptance/rejection decision) and confidence in the correctness of the decision (indicated on a 10-point rating scale).

Materials

The experiment involved 16 one-model syllogisms (half in the AB-BC figure and half in the BA-CB figure). Each figure also involved an equal number of valid and invalid syllogisms and an equal number of traditional affirmative and double negated quantifiers in the premises. According to Roberts’ (2005) notation, the premises were in the moods A[aa] A[aa] and N[an] N[an], for example, *All A are B, All B are C* versus *No A are not B, No B are not C.* Conclusions either matched the premises (i.e., both premises and conclusions were traditional affirmative or both were double negated) or were not matched with the premises (i.e., the traditional affirmative premises were presented with double negated conclusions or double negated premises were presented with traditional affirmative conclusions).

Conflict problems were designated as being those where the use of analytic and heuristic matching strategies were in competition, that is, valid/non-matching problems (e.g., *All A are B, All B are C, Therefore No A are not C*) and invalid/matching problems (e.g., *No A are not B, No B are not C, Therefore No C are not A*). Non-conflict problems were designated as being those where the use of analytic and heuristic matching strategies yielded the same response, that is, valid matching and invalid non-matching problems. The content of all syllogisms involved arbitrary combinations of professions and pastimes. Such contents were systematically rotated through the different problem forms. Authorware 6.5 running on Windows PCs was used to present instructions and problems and to record accept/reject responses, response times and confidence ratings.

Procedure

Participants were presented with 16 syllogisms, one at a time, and were asked to assume that the premises were true. For each syllogism there were three masked statements labelled 'Premise 1', 'Premise 2' and 'Conclusion'. Participants clicked the mouse on a masked area to reveal the statement underneath; each statement was revealed only as long as the mouse pointer remained in that area. As soon as the mouse pointer moved to another location the previously viewed statement was obscured again. Participants were asked to accept conclusions that necessarily followed from the premises (selecting the 'yes' response option) and to reject those which did not (selecting the 'no' response option). They could revisit each masked area as many times as they wished. After each response, participants were asked to record their confidence regarding the response that they had given on a Likert scale that ranged from 1 to 10, with a score of 10 indicating the highest level of confidence.

Results and discussion

A threshold alpha level of .05 was set for all reported analyses, unless otherwise specified.

Acceptance rates

Descriptive statistics for acceptance rate data (Table 1) were analysed using a repeated measures analysis of variance. This revealed a main effect of matching status, $F(1, 84) = 5.95$, $MSE = .10$, $p = .017$, $\eta_p^2 = .07$, with matching conclusions being endorsed more than non-matching ones, indicating a bias toward accepting matching conclusions irrespective of their logical status. There was also a main effect of validity, $F(1, 84) = 28.01$, $MSE = .25$, $p < .001$, $\eta_p^2 = .25$, with more valid conclusions accepted than invalid conclusions, providing some evidence of deductive competence, although acceptance rates for invalid conclusions were still above 50%. In addition, there was a main effect of figure, with more conclusions accepted in the easier AB-BC figure than in the harder BA-CB figure, $F(1, 84) = 11.94$, $MSE = .07$, $p = .001$, $\eta_p^2 = .12$.

Please insert Table 1 about here

The analysis also revealed the presence of a significant three-way interaction between matching status, validity and figure, $F(1, 84) = 5.17$, $MSE = .07$, $p = .026$, $\eta_p^2 = .06$. Post hoc tests (Šidák adjusted threshold alpha level = .004) provided some support for Gilhooly's (2005) claim that participants will employ a more superficial reasoning strategy with complex problems. This was because valid but non-matching conclusions in the difficult BA-CB figure were rejected more frequently than any other valid conclusions (all p values < .001). However, there were no significant differences in acceptance rates for invalid conclusions across the two figures or across matching status.

Response times

Although the primary focus of this experiment was not on the contrast between double negated and traditional affirmative premises, it nevertheless seemed prudent to conduct a response time analysis to determine whether the data replicated the essential slowing of response times for double negated premises, as previously observed by Stupple and Waterhouse (2009). A paired samples t-test successfully replicated the double negation effect, $t(84) = 6.89$, $p < .001$, $d = .59$, showing that participants took significantly longer to respond to problems with double negated premises compared to problems with traditional affirmative premises.

Response time data were skewed and kurtosed. A log transformation was conducted to correct for these violations from a normal distribution. Inferential analyses were subsequently pursued using the transformed data. Descriptive statistics for both the transformed and natural response-time data are reported in Table 2. A repeated measures ANOVA revealed a main effect of matching status, $F(1, 84) = 4.32$, $MSE = .03$, $p = .041$, $\eta_p^2 = .05$, such that non-matching conclusions gave rise to increased response times relative to matching ones. This main effect links with the interaction between conclusion surface features and premise surface features observed by Stupple and Waterhouse (2009). They demonstrated that there were inflated response times for problems with traditional affirmative premises in conjunction with double negated conclusions and for problems with double negated premises in conjunction with traditional affirmative conclusions. In contrast, there were reduced response times for those problems where a matching response was possible. Moreover, the effect of validity demonstrated by Stupple and Waterhouse (2009) was also replicated, $F(1, 84) = 13.56$, $MSE = .03$, $p < .001$, $\eta_p^2 = .14$, such that valid conclusions had shorter response times than invalid conclusions (see also Stupple & Ball 2007; 2008; Thompson et al., 2003).

Please insert Table 2 about here

There was also a significant main effect of figure, $F(1, 84) = 29.06$, $MSE = .03$, $p < .001$, $\eta_p^2 = .26$, with problems in the more difficult BA-CB figure showing inflated response times relative to problems in the easier AB-BC figure. These data further replicate evidence from Stupple and Ball (2007; see also Espino et al., 2000) that figural complexity can have a powerful effect on processing times in conclusion evaluation tasks. Such observations are also consistent with accounts that posit the existence of an increased processing demand for the BA-CB syllogistic figure (e.g., Bara et al. 2001; Johnson-Laird & Bara, 1984). Overall, this figural effect on processing times validates the use of the figural complexity manipulation as a basis for making the theoretically-driven predictions discussed previously concerning the likely impact of problem complexity on acceptance rates.

The analysis additionally revealed the presence of a significant two-way interaction between matching status and validity, $F(1, 84) = 6.75$, $MSE = .02$, $p = .011$, $\eta_p^2 = .07$, whereby non-matching problems engendered inflated response times relative to matching problems, with the caveat that the pattern of results was moderated by the validity of conclusions. Post hoc tests (Šidák adjusted threshold alpha level = .0127) showed significantly increased response times for invalid matching conflict problems relative to valid matching non-conflict problems ($p < .001$), and significantly increased response times for valid non-matching conflict problems relative to valid matching non-conflict problems ($p = .001$). This response time pattern shows similarity to that displayed in the belief bias paradigm, in which the problems exhibiting the most inflated response times are invalid-believable conflict items, where conclusion believability must be inhibited while analytic processing occurs (see Thompson et al., 2003, for the first reported observation of this effect, and Stupple & Ball, 2008, for a replication). It could be argued that increased response times for the invalid matching conflict problems in the present experiment likewise arose from the effort made to inhibit a heuristically-based matching response while analytic processes were applied in an attempt to determine the logical status of the conclusion.

Finally, matching status, validity and figure interacted reliably, $F(1, 84) = 15.65$, $MSE = .03$, $p < .001$, $\eta_p^2 = .16$. Post hoc tests (Šidák adjusted threshold alpha level = .004) indicated that invalid matching conflict problems had longer response times in the harder BA-CB figure than in the easier AB-BC figure ($p < .001$). Similarly, valid non-matching conflict problems had longer response times in figure BA-CB than in figure AB-BC ($p < .001$). It was, moreover, demonstrated that valid matching non-conflict problems in figure BA-CB had shorter response times than invalid matching conflict problems in this figure ($p < .001$). Finally, it was observed that valid non-matching conflict problems had longer response times than valid matching non-conflict problems, but only in the figure BA-CB ($p < .001$).

Overall, these response time analyses indicate that the stand-out data-point relates to the invalid matching problems in the more complex BA-CB figure, which invoke more processing effort than any other items. The acceptance rate data (Table 1) also indicate that the invalid matching problems in the BA-CB figure are more likely to lead to conclusion rejection than other invalid items, although as noted above this effect was not reliable. Nevertheless, there is a suggestion in the descriptive data that the increased processing effort devoted to these problems may be functional in promoting a greater degree of normatively correct conclusion rejection.

Confidence judgements

Descriptive statistics for confidence ratings are reported in Table 3. A repeated measures ANOVA revealed a significant main effect of validity, $F(1, 84) = 11.38$, $MSE = 1.20$, $p = .001$, $\eta_p^2 =.12$, with greater confidence being shown with valid than invalid problems (cf. Quayle & Ball, 2000). There was also a significant main effect of figure, $F(1, 84) = 6.12$, $MSE = 1.02$, $p < .001$, $\eta_p^2 = .07$, with the harder BA-CB problems showing reduced confidence levels relative to the easier AB-BC problems. The latter effect suggests a degree of metacognitive awareness of the increased complexity of the BA-CB problems. This reduced confidence in responses appears to be associated with an increased tendency to reject conclusions to BA-CB problems and for participants also to exhibit longer response times for these items. The effects of both figure and invalidity support the metacognitive uncertainty account (Quayle & Ball, 2000), which predicts that reduced confidence will be induced by invalid problems and by figural complexity (see also Thompson, 2009, for a dual-process perspective on metacognition in reasoning).

Please insert Table 3 about here

Finally, there was an interaction between matching status and validity, $F(1, 84) = 4.02$, $MSE = .60$, $p = .048$, $\eta_p^2 = .05$. Post hoc tests (Šidák adjusted threshold alpha level = .0127) confirmed that confidence was lower for problems with invalid conclusions. However, this effect was only reliable when comparing invalid matching conflict problems to valid matching non-conflict problems ($p < .001$), with there being no significant effect when comparing invalid non-matching non-conflict problems to valid non-matching conflict problems ($p = .09$). As predicted, however, there was also a reduction in confidence for valid non-matching conflict problems compared to valid matching non-conflict problems ($p = .007$, one-tailed). These data, coupled with the acceptance rate findings, are consistent with Prowse Turner and Thompson's (2009) prediction that confidence ratings should decline for conflict problems (see also De Neys et al., 2011). The data also provide some support for De Neys et al.'s (2011) prediction that confidence will be particularly low for participants who give heuristic responses that contradict the normatively sanctioned answer (which is more likely to arise for conflict problems in general). To examine this issue a supplementary individual differences analysis is reported below that focused specifically on the extent to which low confidence levels on conflict problems was associated with increased non-normative responses to those problems.

Individual differences in confidence ratings

De Neys et al. (2011) hypothesised that there would be a contrast in confidence ratings between conflict problems (often responded to incorrectly) and non-conflict problems (often responded to correctly), which would indicate that people have some awareness of the incorrect nature of their responses to conflict items. Across three experiments De Neys et al. showed this predicted confidence reduction for conflict relative to non-conflict items (around 10% for adult participants). Our approach to analysing the association between confidence and reasoning involved first coding participants in terms of the frequency with which they responded to conflict problems in line with heuristics. Such responses contradict the normatively sanctioned evaluation (i.e., heuristic responses involve endorsing invalid matching conclusions and rejecting valid non-matching conclusions). Mean confidence ratings for each conflict problem type were then entered into a multiple regression as predictor variables to test whether confidence was associated with the overall frequency of non-normative responding to conflict items. The regression model was reliable, R = .39, adjusted R^2 = .13, $F(2, 83)$ = 7.03, p = .002, and standardised regression coefficients for each problem type indicated that reduced confidence for valid non-matching problems was associated with increased heuristic responding to conflict items, standardised Beta = .63, $t(83)$ = -3.60, p = .001. This finding supports De Neys et al.'s (2011) prediction and suggests that participants have metacognitive awareness of a violation of normativity when responding heuristically to conflict items. In the case of invalid matching conflict problems, however, the situation was more complex. Contrary to De Neys et al.'s prediction, for these items it was *increased* confidence that was associated with increases in non-normative responding to conflict items, standardised Beta = .40, $t(83)$ = 2.29, p = .025.

A plausible explanation for these confidence findings can be derived from Hardman and Payne (1995; see also Quayle & Ball, 2000), who argued that the presented syllogistic conclusion is central to the mental model construction process in conclusion evaluation tasks. For invalid matching problems the presented conclusion is *possible* given the premises (i.e., it is supported by the initial mental model) but it is not *necessary* (i.e., disconfirming models exist). Thus, for participants seeking to confirm a matching response to invalid matching problems the initial model offers support for accepting the heuristic response, thereby boosting confidence in an incorrect evaluation. In contrast, for valid non-matching conclusions, attempts to find a disconfirming model to the valid conclusion will be in vain (since no such counterexample model exists), so any rejection of a valid non-matching conclusion would be on a heuristic basis, giving rise to reduction in confidence akin to that reported by De Neys et al. (2011).

We also note that confidence differences that arise in reasoning may depend upon whether the responses given involve accepting or rejecting presented conclusions. For example, Thompson, Prowse Turner, and Pennycook (2011) found evidence for reduced 'feelings of rightness' (a measure closely associated with confidence) when participants *rejected* presented conclusions in a belief bias paradigm. This effect persisted even when variance due to the validity and belief status of conclusions was partialled out of the analysis. Thus, it may simply be the case that people express more confidence when responding 'yes' than when responding 'no'. To explore the idea that confidence ratings might polarise according to response type we collapsed our data across participants and analysed 'yes' versus 'no' responses to valid versus invalid items as independent measures. Since the confidence data violated ANOVA assumptions of equal sample sizes and homogeneity of variance we applied a transformation that involved squaring all scores. Although this transformation increased variances it did so without raising skew or kurtosis to problematic levels and it also ensured that variances were homogeneous across conditions (i.e., Levene's test was no longer significant). We also decided to set a highly stringent threshold alpha of .005 for the ANOVA because the smaller sample of responses for conclusion rejections had the greatest within group variance, which can inflate F-ratios.

Confidence ratings demonstrated a significant main effect of response type, with conclusion acceptance associated with greater confidence than conclusion rejection, $F(1, 1352) = 61.48$, $MSE = 827.8$, $p < .001$, $\eta^2 = .04$, thereby supporting Thompson et al.'s (2011) observations. However, there was also a reliable interaction between validity and response type, $F(1, 1352) = 14.49$, $MSE = 827.8$, $p < .001$, $\eta^2 = .01$, in line with predictions that non-normative responses would be associated with decreased confidence. Follow-up analyses indicated that reject responses were associated with lower confidence for valid problems ($M = 7.01$, $SE = 1.60$) than for invalid problems ($M = 7.50$, $SE = 1.40$), $t(407) = 2.21$, $p = .028$, and accept responses were associated with lower confidence ratings for invalid problems ($M = 7.97$, $SE = 1.20$) than for valid problems ($M = 8.34$, SE = 1.10), $t(949) = 3.31$, $p = .001$. No other interactions or effects were significant (all ps > .07). Overall, we suggest that these confidence data are consistent with the predictions of De Neys et al. (2011; see also De Neys, 2012), in that participants appear to have a degree of metacognitive uncertainty when contradicting a normative response.

Individual differences in response times

We conducted a further multiple regression analysis to examine the extent to which the response times for the differing problem types as predictor variables were associated with normative responding as measured by a logic index (total acceptance rates of valid problems minus total acceptance rates of invalid problems). This analysis emulates that employed by Stupple et al. (2011) for examining belief bias data. The regression model that included all four problem types was highly reliable, $R = .50$, adjusted $R^2 = .21$, $F(4, 83) = 6.56$, $p < .001$. Standardised regression coefficients for each problem type indicated that an increase in response times for invalid matching problems was associated with an increase in the overall logic index, standardised Beta = .47, $t(83) = 3.09$, $p = .003$. Moreover, an increase in response times for valid non-matching problems was associated with a *decrease* in the overall logic index, standardised Beta = -.34, $t(83) = -2.45$, $p = .016$. The response times for the valid matching and the invalid non-matching problem types did not make a significant independent contribution toward accounting for the variance in the logic index: valid matching, standardised Beta = .05, $t(83) = .39$, $p = .70$; invalid non-matching, standardised Beta = .22, $t(83) = 1.65$, $p = .10$.

These data are broadly consistent with the explanation proposed above in relation to the individual differences analysis of the confidence data. If participants seek a disconfirming model for a valid non-matching problem then they will take longer to respond (not least because there is no disconfirming model to find), but they will also exhibit a reduction in normative responding whenever they fall back on a heuristically-based reject response. If participants seek a confirmatory model for an invalid matching problem then such a model can readily be discovered. Participants would respond with a non-normative accept decision unless they are able to: (1) inhibit this confirmation bias; and (2) search for a model of the premises that falsifies the conclusion. The latter processes take analytic effort and additional time. In other words, those individuals who respond normatively to conflict problems often appear to have both a stronger sense of the need to search for possible disconfirmatory models as well as a greater capacity to discover such models if they exist. These regression findings show much similarity to the results from belief bias research, which has shown that increased response times to invalid-believable problems are predictive of overall normative responding (Stupple et al., 2011). At the same time, we acknowledge that our data present an intriguing contrast to belief-bias findings, with the observation that the increase in response times for valid non-matching problems is associated with a *decrease* in normative responding. Nonetheless, we view our findings as being explicable from a dual-process framework in which the superficial matching status of conclusions – rather than the belief status of conclusions – is a source of conflict in relation to conclusion validity.

We also pursued a further analysis to examine the issue of whether response latencies polarise according to response type (‘yes’ versus ‘no’ responses), as was the case with the confidence data. Again, we collapsed the data across participants and analysed ‘yes’ versus ‘no’ responses to valid versus invalid items as independent measures. We corrected violations of homogeneity of variance using a log transformation and set a stringent alpha of .005 for the ANOVA. The analysis revealed a significant effect of response type, with rejections (M = 21.5, SE = 1.04) taking longer than acceptances (M = 16.26, SE = 1.02), $F(1, 1352) = 48.54$, MSE = .08, $p < .001$. These results replicate the findings of Thompson et al.’s (2011). No other effects or interactions were significant (all ps > .08), suggesting that while confidence dips for non-normative responses, this does not inspire additional processing time.

General discussion

The present study provides evidence for the influence of dual reasoning processes on syllogistic inference when surface features of presented syllogisms come into conflict with conclusion validity. These surface features are not related to beliefs, as in the traditional belief bias paradigm that examines logic/belief conflicts in reasoning. Instead, they concern whether premises and conclusions match or mismatch in terms of the presence of double negated quantifiers (Roberts, 2005) or standard affirmative quantifiers. The difference between the belief bias paradigm and the present paradigm is important, since in the present study it is not people’s ability to ‘decontextualise’ the content of problems from real-world knowledge that is in question, but rather their ability to divorce their analytic processing from the biasing influence of the surface features of the quantifiers within the problem. Despite the differences in the nature of the response conflicts that arose in the present paradigm compared to the traditional belief bias paradigm, we have demonstrated some important parallels with findings arising in the more established literature on belief bias. One key correspondence concerns the observation that conflict problems in both paradigms show inflated response times relative to non-conflict problems (Ball et al., 2006; Stupple & Ball, 2008; Thompson et al., 2003), which is fully in line with dual-process predictions that derive from either a parallel process perspective (e.g., Sloman, 2002; Stupple & Ball, 2008) or a default-interventionist perspective such as the selective processing model espoused by Evans (e.g., 2000; see also Stupple et al., 2011).

At a more detailed level of analysis the present chronometric data also reveal some important associations with normative response rates that appear to align reasonably well with specific concepts that are embodied within extant dual-process models such as the selective processing model (Evans, 2000; Stupple et al., 2011). One such concept concerns the way in which biases that arise from *either* heuristic processing *or* the selective nature of analytic processing are associated with reduced normative responding, with participants who are able to inhibit or avoid such biases being more likely to respond normatively (Stupple et al., 2011). The present study suggests that such normative responding can be at the cost of additional processing time with invalid matching problems, which is consistent with Stupple et al.'s belief bias data for invalid believable problems.

However, an interesting contrast arises when comparing the present findings with belief bias data, which relates to valid-unbelievable conflict problems in belief bias studies and valid non-matching conflict problems as examined here. In belief bias research the response times for valid-unbelievable conflict problems are not generally found to be associated with overall normative responding (Stupple et al., 2011), whereas here the response times for valid non-matching conflict problems *were* associated with normative responding, but in the opposite direction to that expected (i.e., it was *faster* responses to these problems that were associated with success in making normative evaluations). We have suggested that this counterintuitive finding could be explicable in terms of the differing model searches arising for valid and invalid problems (cf. Hardman & Payne, 1995). Alternatively, it could be proposed that a sub-set of participants applies a 'negation-elimination' strategy (cf. Rips, 1994), which allows them rapidly to avoid the biasing influence of the surface features of problems. Yet another interesting and parsimonious explanation suggested by a reviewer is couched in parallel-process terms, with the idea being that an awareness of the valid/non-matching conflict arises *later* than the response that is cued by the logic of the problem. The longer participants take to respond then the more likely they will be to detect the conflict, which could engender incorrect conclusion rejection.

We acknowledge, however, that all of these latter proposals are post hoc and reflect an attempt to explain an anomalous result that ran counter to expectations. Again, this underlines the fact that although the general tenets of dual-process models often find support in the literature there are occasional findings that are more challenging to accommodate. An important direction for future research in relation to these issue concerns the need to clarify the extent to which tendencies toward normative responding are a function of available processing recourses, dispositional factors or a combination of both. In terms of processing resources, it could be that what is crucial for the effective analytic processing of conflict problems is the availability of an increased degree of raw computational power that would come from having an elevated working memory capacity. Thus, the cognitively demanding act of inhibiting a temptingly frugal heuristic response while searching exhaustively for alternative models of the premises might only be possible for individuals with the largest working memory capacities (see Copeland & Radvansky, 2004, for evidence linking working memory span and performance on belief-neutral syllogisms). On the other hand, it may be that it is dispositions to engage in critical thinking that lie at the heart of the generation of normative responses to conflict items. Such dispositions are detectable through tests that assess the degree to which people are inclined to think critically about problems as opposed to being driven by their intuitions. Prime examples of such tests are the Cognitive Reflection Test (Frederick, 2005), the Actively Open Minded Thinking Test (Stanovich & West, 1997) and the Rational-Experiential Inventory (Epstein, 1994). Previous research has shown a link between such thinking dispositions and normative responding on reasoning tasks (e.g., Macpherson & Stanovich, 2007; Stanovich, West, & Toplak, 2010). The emerging picture in the literature seems to be that *both* cognitive resource factors and dispositional factors are important in predicting individual differences in normative responding (e.g., Stanovich, 2009; see also De Neys, 2006b; Torrens, Thompson, & Cramer 1999), but further work is clearly required to investigate more fully the interplay between such factors.

The findings in the present study concerning conclusion endorsement rates also provide some support for the notion that increased problem complexity arising from a figural manipulation results in greater recourse to a simplified reasoning strategy, as proposed by Gilhooly (2005; see also Quayle & Ball, 2000). However, the data reveal that this complexity effect is somewhat nuanced and only arises for problems that have valid conclusions. Turning to the confidence data, we found evidence for Shynkaruk and Thompson's (2006) prediction that confidence is to some extent determined by non-logical factors. For example, we found evidence for both figural complexity and conclusion validity impacting confidence levels. In their experiments, however, participants were more confident about belief-oriented problems than belief-neutral problems. Although specific effects for conflict problems were not examined by Shynkaruk and Thompson, in related research clear evidence of decreased confidence was observed for conflict items (see Prowse Turner & Thompson, 2009); such findings are supported in the present study, where a majority of comparisons revealed reduced confidence for conflict items relative to non-conflict items.

Of additional interest is the observation in the present study that participants' confidence judgements were indicative of some level of awareness of normative considerations, even when they responded non-normatively (cf. De Neys et al., 2011). Stanovich (2009) presents a taxonomy of thinking errors, characterised at the highest level in terms of 'cognitive miser' and 'mindware gap' categories. We would argue that the cognitive miser category is implicated in our findings since many participants appear either to default to a heuristic response or to fail successfully to override such a response. Moreover, the 'override failure' explanation gains further support from the evidence that participants appear to possess the necessary mindware to detect a conflict between heuristic and analytic processes, as confidence judgements and response times revealed an appreciation of the normative status of presented conclusions even when they made a non-normative evaluation (cf. De Neys, 2012; Morsanyi & Handley, 2012; Stupple et al., 2011). Nonetheless, it could also be argued that some participants were able to avoid falling foul of a matching bias as the result of superior mindware. For example, the possession of a 'negation-elimination' strategy could allow some participants to avoid confronting and having to resolve heuristic/analytic conflicts.

We also note that although a dual-process perspective has been presented throughout this article as an explanatory framework to derive predictions and to explain the data, it might nevertheless be possible for critics of dual-process theory (e.g., Dube et al., 2010; Keren & Schul, 2009) to develop a single-process model to account for our findings. We suggest, however, that developing such a model would be challenging given the sheer range of evidence obtained in the present study, deriving as it does from multiple converging measures. Our present data, moreover, appear to align very closely with other recent evidence in the literature arising from several studies of heuristic/analytic conflict detection in reasoning, judgement and decision making (e.g., Bonner & Newell, 2010; De Neys et al., 2008, 2011; Stupple et al., 2011; Thompson, Morley, & Newstead, 2011). Again, this latter evidence ranges across data that include acceptance rates, response times, metacognitive confidence judgements and individual differences measures. It is certainly to the credit of dual-process theories that they are not only capable of explaining much of the data that arise from such a wide range of measures but that they also typically predict from the outset the various effects that are subsequently observed.

References

Ball, L. J. (2010). The dynamics of reasoning: Chronometric analysis and dual-process theories. In K. I. Manktelow, D. E. Over, & S. Elqayam (Eds.), *The science of reason: A Festschrift for Jonathan St. B. T Evans* (pp. 283-307). Hove: Psychology Press.

Ball, L. J., Phillips, P., Wade, C. N., & Quayle, J. D. (2006). Effects of belief and logic on syllogistic reasoning: Eye-movement evidence for selective processing models. *Experimental Psychology*, *53,* 77-86.

Bara B. G., Bucciarelli, M., & Lombardo, V. (2001). Mental model theory of deduction: A unified computational approach. *Cognitive Science, 25*, 839-901.

Bonner, C., & Newell, B. R. (2010). In conflict with ourselves? An investigation of heuristic and analytic processes in decision making. *Memory & Cognition, 38*, 186-196.

Bucciarelli, M., & Johnson-Laird, P. N. (1999). Strategies in syllogistic reasoning *Cognitive Science*, *23,* 247-303.

Chater, N., & Oaksford, M. (1999). The probability heuristics model of syllogistic reasoning. *Cognitive Psychology, 38,* 191-258.

Copeland, D., & Radvansky, G. (2004). Working memory and syllogistic reasoning. *Quarterly Journal of Experimental Psychology*, *57A*, 1437-1457.

De Neys, W. (2006a). Automatic–heuristic and executive–analytic processing during reasoning: Chronometric and dual-task considerations. *Quarterly Journal of Experimental Psychology, 59* 1070-1100.

De Neys, W. (2006b). Dual processing in reasoning: Two systems but one reasoner. *Psychological Science, 17*, 428-433.

De Neys, W. (2012). Bias and conflict: A case for logical intuitions. *Perspectives on Psychological Science, 7,* 28-38.

De Neys, W., Cromheeke, S., & Osman, M. (2011). Biased but in doubt: Conflict and decision confidence. *PLoS ONE, 6*, e15954.

De Neys, W., & Franssens, S. (2009). Belief inhibition during thinking: Not always winning but at least taking part. *Cognition, 113,* 45-61.

De Neys, W., Moyens, E., & Vansteenwegen, D. (2010). Feeling we're biased: Autonomic arousal and reasoning conflict. *Cognitive, Affective, and Behavioral Neuroscience, 10,* 208-216.

De Neys, W., Vartanian, O., & Goel, V. (2008). Smarter than we think: When our brains detect that we are biased. *Psychological Science, 19*, 483-489.

Dowty, D. (1994). The role of negative polarity and concord marking in natural language reasoning. In M. Harvey & L. Santelmann (Eds.), *Proceedings of Semantics and Linguistic Theory IV* (pp. 114-144). Ithaca. NY: Cornell University.

Dube, C., Rotello, C. M., & Heit, E. (2010). Assessing the belief bias effect with ROCs: It's a response bias effect. *Psychological Review, 117,* 831-863.

Elqayam, S. (2009). Models of dependence and independence: A two-dimensional architecture of dual processing. *Thinking & Reasoning, 15,* 377–387.

Elqayam, S., & Evans, J. St. B. T. (2011). Subtracting "ought" from "is": Descriptivism versus normativism in the study of human thinking. *Behavioral & Brain Sciences, 34*, 233-248.

Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *American Psychologist, 49*, 709-724.

Espino, O., Santamaria, C., & Garcia-Madruga, J. A. (2000). Figure and difficulty in syllogistic reasoning. *Current Psychology of Cognition 19,* 417-428.

Evans, J. St. B. T. (1989). *Bias in human reasoning: Causes and consequences*. Hove: Erlbaum.

Evans, J. St. B. T. (2000). Thinking and believing. In J. Garcìa-Madruga, N. Carriedo, & M. J. González-Labra (Eds.), *Mental models in reasoning* (pp. 41-56). Madrid, Spain: UNED.

Evans, J. St. B. T. (2007). On the resolution of conflict in dual process theories of reasoning. *Thinking & Reasoning, 13,* 321-339.

Evans, J. St. B. T. (2009). How many dual process theories do we need? One, two, or many? In J. St. B. T. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 35-54). Oxford: Oxford University Press.

Evans, J. St. B. T. (2011). Dual-process theories of reasoning: Contemporary issues and developmental applications. *Developmental Review*, *31,* 86-102.

Evans, J. St. B. T., & Ball, L. J. (2010). Do people reason on the Wason selection task? A new look at the data of Ball et al. (2003). *Quarterly Journal of Experimental Psychology, 63,* 434-441.

Evans, J. St. B. T., Barston, J. L., & Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory & Cognition*, *11,* 295-306.

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review, 103,* 650-669.

Gilhooly, K. J. (2005). Working memory and strategies in reasoning. In M. J. Roberts & E. J. Newton (Eds.), *Methods of thought: Individual differences in reasoning strategies.* Hove: Psychology Press.

Hardman, D. K., & Payne, S. J. (1995). Problem difficulty and response format in syllogistic reasoning. *Quarterly Journal of Experimental Psychology, 48A,* 945-975.

Johnson-Laird, P. N. (1983). *Mental models: Toward a cognitive science of language, inference and consciousness*. Harvard: Harvard University Press.

Johnson-Laird, P.N., & Bara, B.G. (1984). Syllogistic inference. *Cognition 16*, 1-61.

Johnson-Laird, P. N., & Byrne, R. M. J. (1989). *Only* reasoning. *Journal of Memory and Language, 28,* 313-330.

Keren, G., & Schul, Y. (2009). Two is not always better than one: A critical evaluation of two-system theories. *Perspectives on Psychological Science, 4,* 533-550.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality & Social Psychology, 77*, 1121-1134.

Luo, J., Liu, X., Stupple, E. J., Zhang, E., Xiao, X., Jia, L., Yang, Q., Li, H., & Zhang, Q. (in press). Cognitive control in belief-laden reasoning during conclusion processing: An ERP study. *International Journal of Psychology.*

Luo, J., Yuan, J. J., Qiu, J., Zhang, Q. L., Zhong, J., & Huai, Z. C. (2008). Neural correlates of the belief-bias effect in syllogistic reasoning: an event-related potential study. *Neuroreport, 19*, 1075-1079.

Macpherson, R., & Stanovich, K. E. (2007). Cognitive ability, thinking dispositions, and instructional set as predictors of critical thinking. *Learning & Individual Differences*, *17,* 115-127.

Morsanyi, K., & Handley, S. J. (2012). Logic feels so good -I like it! Evidence for intuitive detection of logicality in syllogistic reasoning. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 38,* 596-616.

Prowse Turner, J. A., & Thompson, V. A. (2009). The role of training, alternative models and logical necessity in determining confidence in syllogistic reasoning. *Thinking & Reasoning, 15*, 69-100.

Quayle, J. D., & Ball, L. J. (2000). Working memory, metacognitive uncertainty, and belief bias in syllogistic reasoning. *Quarterly Journal of Experimental Psychology, 53A*, 1202-1223.

Rips, L. (1994). *The psychology of proof*. Cambridge, MA: MIT Press.

Roberts, M. J. (2005). Expanding the universe of categorical syllogisms: A challenge for reasoning researchers. *Behavior Research Methods, 37*, 560-580.

Shynkaruk, J. M., & Thompson, V. A. (2006). Confidence and accuracy in deductive reasoning. *Memory & Cognition, 34,* 619-632.

Sloman, S. A. (2002). Two systems of reasoning. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 379-398). Cambridge: Cambridge University Press.

Stanovich, K. E. (2004). *The robot's rebellion: Finding meaning the age of Darwin.* Chicago: Chicago University Press.

Stanovich, K. E. (2009). Distinguishing the reflective, algorithmic and autonomous minds: Is it time for a tri-process theory? In J. St. B. T. Evans, & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 55-88). Oxford: Oxford University Press.

Stanovich, K.E., & West, R. F. (1997). Reasoning independently of prior belief and individual differences in actively open minded thinking. *Journal of Educational Psychology*, 89, 342-357.

Stanovich, K. E., West, R. F., & Toplak, M. E. (2011). Individual differences as essential components of heuristics and biases research. In K. I. Manktelow, D. E. Over, & S. Elqayam

(Eds.), *The science of reason: A festschrift for Jonathan St. B. T. Evans* (pp. 335-396). Hove: Psychology Press.

Stupple, E. J. N., & Ball, L. J. (2007). Figural effects in a syllogistic evaluation paradigm: An inspection-time analysis. *Experimental Psychology, 54,* 120-127.

Stupple, E. J. N., & Ball, L. J. (2008). Belief-logic conflict resolution in syllogistic reasoning: Inspection-time evidence for a parallel-process model. *Thinking & Reasoning, 14,* 168-181.

Stupple, E. J. N., & Ball, L. J. (2011). Normative benchmarks are useful for studying individual differences in reasoning. *Behavioral & Brain Sciences, 34*, 270-271.

Stupple, E .J. N., Ball, L. J., Evans, J. St. B. T., & Kamal-Smith, E. N. (2011). When logic and belief collide: Individual differences in reasoning times support a selective processing model. *Journal of Cognitive Psychology, 23*, 931-941.

Stupple, E. J. N., & Waterhouse E. F. (2009). Negations in syllogistic reasoning: evidence for a heuristic-analytic conflict. *Quarterly Journal of Experimental Psychology, 62*, 1533-1541.

Thompson, V. A. (2009). Dual process theories: A metacognitive perspective. In J. St. B. T. Evans, & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 171-195). Oxford: Oxford University Press.

Thompson, V. A., Morley, N. J., & Newstead, S. E. (2011). Methodological and theoretical issues in belief-bias: Implications for dual process theories. In K. I. Manktelow, D. E. Over, & S. Elqayam (Eds.), *The Science of Reason: A Festschrift for Jonathan St. B. T Evans* (pp. 309-338). Hove: Psychology Press.

Thompson, V. A., Prowse Turner, J. A., & Pennycook, G. (2011). Intuition, reason, and metacognition. *Cognitive Psychology, 63,* 107-140.

Thompson, V. A., Striemer, C. L., Reikoff, R., Gunter, R. W., & Campbell, J. D. (2003). Syllogistic reasoning time: Disconfirmation disconfirmed. *Psychonomic Bulletin & Review, 10*, 184-189.

Torrens, D., Thompson, V. A., & Cramer, K. M. (1999). Individual differences and the belief bias effect: Mental models, logical necessity, and abstract reasoning. *Thinking & Reasoning* 5, 1-28.

Wetherick, N. E., & Gilhooly, G. J. (1995). 'Atmosphere', matching, and logic in syllogistic reasoning. *Current Psychology 14*. 169-178.

Table 1.

Mean Percentage Acceptance Rates as a Function of Matching Status, Validity and Figure

	Matching			Non-matching		
	AB-BC	BA-CB	Total	AB-BC	BA-CB	Total
Valid	83.9	83.3	83.6	83.3	69.1	76.2
Invalid	66.1	57.1	61.6	59.5	54.8	57.2
Total	75.0	70.2	72.6	71.4	61.9	66.7

Note: Standard errors ranged from 3.1 to 4.7.

Table 2.

Mean Response Times (in Seconds) as a Function of Matching Status, Validity and Figure for Natural Data (Panel a) and for Transformed Data Converted Back into Original Units (Panel b)

(a)

	Matching			Non-matching		
	AB-BC	BA-CB	Total	AB-BC	BA-CB	Total
Valid	19.0	20.3	19.7	19.4	24.5	22.0
Invalid	19.4	27.4	23.4	22.2	24.3	23.3
Total	19.2	23.85	21.6	20.8	24.4	22.7

Note: Standard errors ranged from 1.07 to 1.59.

(b)

	Matching			Non-matching		
	AB-BC	BA-CB	Total	AB-BC	BA-CB	Total
Valid	16.6	17.0	16.8	17.0	21.2	19.1
Invalid	17.0	23.4	20.2	19.3	20.3	19.8
Total	16.8	20.2	18.5	18.2	20.8	19.5

Note: Standard errors ranged .02 to .03 for transformed data.

Table 3.

Mean Confidence Ratings (Maximum Score = 10) as a function of Matching status, Validity and Figure

	Matching			Non-matching		
	AB-BC	BA-CB	Total	AB-BC	BA-CB	Total
Valid	8.00	7.88	7.94	7.78	7.61	7.70
Invalid	7.68	7.25	7.47	7.49	7.44	7.47
Total	7.84	7.57	7.71	7.64	7.53	7.58

Note: Standard errors ranged from .19 to .22.